DOCUMENT RESUME

ED 426 086 TM 029 314

AUTHOR Henson, Robin K.
TITLE ANCOVA with Intact Groups: Don't Do It!
PUB DATE 1998-11-00
NOTE 30p.; Paper presented at the Annual Meeting of the Mid-South Educational Research Association (New Orleans, LA, November 4-6, 1998).
PUB TYPE Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Analysis of Covariance; *Educational Research; *Groups; Heuristics; Hypothesis Testing

ABSTRACT

This paper explains how analysis of covariance (ANCOVA) and related statistical corrections work and discusses difficulties with the use of these corrections under certain circumstances. ANCOVA is essentially a regression of a covariate variable on the dependent variable from the entire sample ignoring group membership, at least, if ANCOVA assumptions are perfectly met. Small heuristic data sets are used to illustrate when ANCOVA can and cannot be used correctly in educational research. ANCOVA generally can be used correctly with randomly assigned groups, but may not be needed in that situation. When groups are not randomly assigned, ANCOVA often cannot be used correctly. It is a paradox that ANCOVA often cannot be used when "correction" is most needed. For ANCOVA, meeting the homogeneity of regression assumption is critical in determining its viability as a statistical tool. An appendix contains a computer program for the ANCOVA heuristic analysis. (Contains four tables, five figures, and eight references.) (Author/SLD)

Running head: ANCOVA with Intact Groups

ANCOVA with Intact Groups: Don't Do It!

Robin K. Henson

Texas A&M University

Paper presented at the annual meeting of the Mid-South Educational Research Association, New Orleans, November 4, 1998.

Abstract

The present paper explains how ANCOVA and related statistical corrections work, and discusses difficulties with the use of these corrections under certain circumstances. Small heuristic data sets are employed to illustrate when ANCOVA can and can not be correctly used in educational research. In the main, ANCOVA can usually be correctly used with randomly assigned groups, but may not be needed here. When groups are not randomly assigned, ANCOVA often can not be correctly employed. Paradoxically, ANCOVA often can not be used when "correction" is most needed.

ANCOVA with Intact Groups: Don't Do It!

True experimental designs are a rarity in educational research. While many researchers yearn for the possibility of testing curriculum, achievement interventions, and other research protocols, the nature of many educational systems do not allow random assignment of students to intervention conditions. True experiments become increasingly rare in a litigious environment populated by overly conservative and intrusive Institutional Review Boards. For example, Welch and Walberg (1974, p. 113) noted that

> Although the need for true experiments on broadly defined populations has long been recognized, there are very few local experiments and no national experiments in curriculum research. For example, among 46 government-sponsored course development projects in science and mathematics, a few relied on teacher reports and classroom visits for evaluation, but only four used true experiments in their evaluation strategies.

Thus, researchers often find themselves at the mercy of established groupings for which they then must develop interventions and statistical analyses. Since true experimental control with random assignment of subjects is often impossible or even unethical, one approach to dealing with intact groups in educational research is to attempt to statistically control for the pre-existing differences between the groups being studied. These statistical procedures are all related to each other, and go by various names, such as analysis of covariance (ANCOVA) and partial correlation. Though many of these statistical controls date back to the beginning of the century, most of the corrections have not enjoyed especially wide use in journal articles. ANCOVA's appearance, for example, in educational literature is limited to about 4% of the recently published research (Elmore & Woehlke, 1988; Goodwin & Goodwin, 1985; Willson, 1980). Thompson (1988, 1994) found that doctoral students use ANCOVA with greater frequency.

Perhaps the lingering use of ANCOVA is due to the mystical promise that ANCOVA is a statistical correction of all pre-treatment problems and that it will provide increased power against Type II error. Such an argument is particularly compelling to doctoral students who find themselves aggressively seeking and even praying for statistically significant results! Unfortunately, ANCOVA has multiple assumptions that must be met before it can be accurately utilized.

The purpose of the present paper is (a) to highlight the important distinction between statistical analysis and methodological design as it relates to the use of ANCOVA, (b) to discuss how ANCOVA purports to make statistical corrections, and (c) to illustrate the necessity of meeting the homogeneity of regression assumption in all uses of ANCOVA. A small heuristic data set will be employed to aid in illustration of this final purpose.

Analysis Versus Design

It is not uncommon for researchers to confuse the concepts of statistical analysis and methodological design. In fact, these concepts represent two related but separate issues in designing and conducting quality research. ANCOVA may be seen as a statistical way for dealing with methodological design flaws.

If a researcher has three classrooms that he or she can study, these pre-existing groupings may prohibit him or her from randomly assigning students to one of three treatment conditions. As such, a researcher is compelled to use the three intact groups as they exist in the real world and assume low levels of sampling error. Obviously, such an assumption is both inappropriate and reflective of poor research. Since the given researcher knows that the design is at great risk for heavy influence due to sampling error, he or she is then compelled to invoke a statistical process, namely ANCOVA, that will serve as a method of equating the "groups or statistically removing from dependent variable variance the effects of a continuously-scaled extraneous variable (or variables) so that treatment effects can be clarified and the probability of obtaining statistically significant results will be increased" (Loftin & Madison, 1991, p. 133).

When accurately applied, ANCOVA actually can fulfill its promise as described by Loftin and Madison (1991). However, when inaccurately applied, one must seriously question the consequences of such an analysis. One possibility lies in researcher ignorance or, even worse, utter disregard for the assumptions in the correct use of ANCOVA. (While the homogeneity of regression assumption is addressed in the present paper, a complete list and discussion of other conditions for ANCOVA may be found in Loftin and Madison (1991).)

However, a second possibility seems just as likely. That is, ANCOVA may be viewed as a statistical method that is part and parcel with the research design. In a researcher's zealous attempts to have an experimental or quasi-experimental design, the use of ANCOVA comes to be seen as a way of making such a design happen. In fact, a design is either experimental or it is not. Statistical analyses employed on the data obtained from the design do not magically transform a study into a true experiment. Such thinking is obviously flawed, but all too common, at least in dissertation research (Thompson, 1994). This thinking (or lack thereof) reflects confusion regarding the cooperative but separate roles of methodological design and statistical analysis.

What does ANCOVA Purportedly Correct?

ANCOVA is essentially a regression of a covariate variable on the dependent variable from the entire sample ignoring group membership, at least if ANCOVA assumptions are perfectly met. The intent of this process is to assign a portion of the variance in the dependent variable that would normally be attributed to error in a regular analysis of variance (ANOVA) to an extraneous covariate variable. This will result in reduced error sum of squares. The independent variable treatment effects, then, can be "clarified" because the researcher has supposedly eliminated some of the sampling error due to a lack of randomization (Loftin & Madison, 1991, p. 133).

After the covariate regression is performed, an ANOVA is performed on the residualized dependent variable, that is, the error scores that remain after assigning part of the dependent variable variance as due to the covariate. The independent variable's (e.g., the treatment way's) F calculated is enhanced because the error variance is less than had the covariate influence not been removed. Figures 1 through 3 illustrate this process using Venn diagrams to represent the total sum of squares (SOS) in the dependent variable. Figure 1 reflects a classical ANOVA analysis with an effect size of 25% (eta^2=.25). Figure 2 demonstrates the first step in an ANCOVA analysis, a regression in which the covariate effect size is 25% (r^2=.25). Figure 3 carries out the ANCOVA analysis by performing an ANOVA between the independent variable and the remaining error scores after the regression in Figure 2. The result is an enhanced likelihood of obtaining statistical significance. The SOS error in Figure 3 is reduced to 66.66% of the total dependent variable variance from 75% in the classical ANOVA in Figure 1.

INSERT FIGURES 1-3 ABOUT HERE.

When applied appropriately, ANCOVA can have such an outcome. However, there are very few situations in which all of the assumptions are met. One condition is that the covariate must be highly correlated with the dependent variable but with no or a very low correlation with the independent variable. This condition is met in Figure 3. In the real world, however, it may be difficult to find meaningful covariate variables that are not correlated with the independent variable, especially when people are not randomly assigned to experimental conditions. If there exists a correlation between the covariate and the independent variable, then the covariate can actually take away from the variance (and subsequent effect size) of the independent variable. Figure 4 illustrates this dynamic.

INSERT FIGURE 4 ABOUT HERE.

ANCOVA's supposed promise of power against Type II error is only fulfilled in a case in which the covariate is uncorrelated with the independent variable. Otherwise the covariate robs the independent variable of variance attributable to it. As Thompson (1994) noted:

> When the covariate is related to the treatment variable, use of the covariance correction will alter the effects attributed to the treatment itself. For example, one might have a very effective intervention that looks completely ineffectual, because the covariate is given credit for the variance that would correctly otherwise be attributed to the treatment variable. Here the ANCOVA correction actually destroys power against Type II error. (p. 27)

Of course, the total effect of the covariate on the independent and dependent variable is a function of its relationship to both of these variables. It is possible that the covariate is highly correlated with both variables, yielding a result similar to that had a covariate not even been used in the analysis. The covariate may reduce error but may also rob the treatment variable proportionately.

The Homogeneity of Regression Assumption

The essential goal of ANCOVA is to equate groups on some variable before analyzing the effects of treatments. This is attempted by performing regression of the covariate on the dependent variable, ignoring group membership. Logically, if one regression equation is to be used for all groups (since group membership is ignored), then logically the same equation must also be representative of each of the groups examined individually, or otherwise the use of the single "pooled" correction equation will distort all the dependent variable scores.

This is the homogeneity of regression assumption. That is, the regression slopes of the covariate and the dependent variable in the individual groups must be the same if the single pooled regression slope can be accurately used with all groups. If the individual regression slopes are notably different, then the pooled regression slope will not effectively represent any of the groups! If this occurs, and the pooled regression equation does not accurately reflect the regression equation in a given group, then the use of ANCOVA will "actually bias the data rather than 'correct' them" (Loftin & Madison, 1991, p. 143).

In efforts to illustrate this assumption, a small data set will be employed for heuristic purposes. In the hypothetical example here, the experimenter is concerned with examining the effects of a teaching effectiveness intervention for special education students. Namely, the researcher wants to determine if a note-taking strategy will positively impact at-risk student achievement. Unfortunately, as with many educational "experiments", the researcher must provide the note-taking intervention to all of the school's intact special education classrooms. The administration of the school is concerned about the possible negative reactions from parents if they deny a potentially helpful intervention to some of the special education students. As such, the experimenter is left without a true control group and decides to use a classroom of mainstream education students as a control. In order to "equate" the groups so they can be compared, the researcher will utilize ANCOVA, which purportedly promises to "control" for pre-existing differences that probably exist between the two classrooms being studied.

In this case, the dependent variable is a standardized achievement test measure (ACHIEVE). The covariate selected by the researcher is a reading and writing achievement measure (READ). This variable was selected because it is intended to "nullify" the pre-existing verbal difference between the special and mainstream education students. This is, after all, a note-taking strategy intervention and obvious differences would probably exist in reading and writing skills between groups. The independent variable used in this ANCOVA shall be the special education treatment versus the mainstream education control groups. Table 1 gives the data for the treatment and controls groups. Scores on the READ and ACHIEVE measures are given as T-scores.

INSERT TABLE 1 ABOUT HERE.

As would be expected, the special education treatment group generally scored lower on the covariate READ measure (M=38.33, SD=6.83) than the treatment group (M=56.67, SD=6.83). Since the intervention deals highly with verbal and reading skills, the researcher then uses this data to "level" the two groups so they can be compared on the

ACHIEVE posttest. In doing so, the researcher performs a standard ANCOVA analysis with results reported in Table 2. The same results can be achieved by performing a regression of the covariate (READ) on the dependent measure (ACHIEVE) and then performing a classical ANOVA on the residualized dependent variable. Table 3 reports results of both of these analyses. The SPSS syntax file used to determine these results can be found in Appendix A.

INSERT TABLES 2-3 ABOUT HERE.

These results illustrate the synonymous nature of an ANCOVA analysis and a regression followed by an ANOVA analysis. Another point to note is the disparity between the ANCOVA effect size (eta^2=45.44%) and the effect size that would have been attained if just a classical ANOVA had been used (eta^2=78.48%). In this case, the covariate clearly was correlated with the independent variable and robbed the treatment effect of some glory when it was residualized out of the dependent variable! Table 4 reports results from a classical ANOVA analysis.

INSERT TABLE 4 ABOUT HERE.

In the regression of the covariate (READ) on the dependent variable (ACHIEVE), a pooled regression equation is created that ignores group membership (Y = -9.158192 + 1.405085 * READ). As noted previously, the homogeneity of regression assumption demands that this pooled equation reasonably represent the separate equations of the individual groups. Specifically, the slopes should be similar. Figure 5 graphically illustrates that this assumption is not met in the present case. After performing a regression of the treatment and "control" groups individually, one discovers slopes (b weights) of .986 and 1.257, respectively.

INSERT FIGURE 5 ABOUT HERE.

One can clearly see from Figure 5 that the pooled regression slope can not reasonably be used to represent both groups. This is a case in which the use of ANCOVA would "bias the data rather than 'correct' them" (Loftin & Madison, 1991, p. 143). In the real world, it is often the case when the b weights are different between groups that random assignment was not utilized. The use of ANCOVA with intact groups lends itself to disparate b weights between such groupings, because the disparate characteristics of such groups also impact the covariate relationships with the dependent variable scores.

Even beyond not meeting this assumption one must question what is left of the ACHIEVE dependent variable after the dramatic impact of the covariate (r^2=96.01%). Thompson (1992, pp. xiii-xiv) addresses this issue of dependent variable interpretation by emphasizing, "Statistical corrections remove parts of the dependent variable, and then analyze whatever's left, even if whatever's left no longer makes any sense. At some point we may no longer know what it is we're analyzing."

When can ANCOVA be Used?

As with most statistical analyses, ANCOVA can be reasonably applied when its assumptions have been met. For ANCOVA, meeting the homogeneity of regression assumption is critical in determining its viability as a statistical tool. Paradoxically, as b weights become more similar, the less need there is to make a statistical correction to equate groups because the groups are being equated by random assignment. It seems, then, that ANCOVA can, at best, serve to subtly refine true experiments that utilize random assignment with a large enough sample size for such assignment to work. At worst, ANCOVA can be misused as some sort of mystical and ineffective/deleterious equating of groups that are, in fact, not equal. This reflects a misuse of the analysis and yields results and conclusions not supported by the data.

References

Elmore, P. B., & Woehlke, P. L. (1988). Statistical methods employed in American Educational Research Journal, Educational Researcher and Review of Educational Research from 1978 to 1987. Educational Researcher, 17(9), 19-20.

Goodwin, L. D., & Goodwin, W. L. (1985). Statistical techniques in AERJ articles, 1979-1983: The preparation of graduate students to read the educational research literature. Educational Researcher, 14(2), 5-11.

Loftin, L. B., & Madison, S. Q. (1991). The extreme dangers of covariance corrections. In B. Thompson (Ed.), Advances in educational research: Substantive findings, methodological developments (Vol. 1, pp. 133-147). Greenwich, CT: JAI Press.

Thompson, B. (1988, November). Common methodology mistakes in dissertations: Improving dissertation quality. Paper presented at the annual meeting of the Mid-South Educational Research Association, Louisville, KY. (ERIC Document Reproduction Service No. ED 301 595)

Thompson, B. (1992). Misuse of ANCOVA and related "statistical control" procedures. Reading Psychology: An International Quarterly, 13, iii-xvii.

Thompson, B. (1994, April). Common methodology mistakes in dissertations, revisited. Paper presented at the annual meeting of the American Educational Research Association, New Orleans. (ERIC Document Reproduction Service No. ED 368 771)

Welch, W. W., & Walberg, H. J. (1974). A course evaluation. In H. J. Walberg (Ed.), Evaluating educational performance: A sourcebook of methods, instruments, and examples (pp. 113-124). Berkeley: McCutchan.

Willson, V. L. (1980). Research techniques in AERJ articles: 1969 to 1978. Educational Researcher, 9(6), 5-10.

Table 1

Heuristic Data for Treatment and Control Students on Reading (READ) and Achievement (ACHIEVE) Measures (n=12)

Student	READ	ACHIEVE
Special Education Students Receiving Intervention		
Jennifer	30	34
Suzanne	30	36
Greg	40	46
Kyle	40	46
Natascha	45	49
Alfred	45	50
M	38.33	43.50
SD	6.83	6.80
Students in Regular Education Without Intervention		
Stephanie	50	60
Bob	50	68
Timothy	55	67
Micah	55	70
Elizabeth	65	80
Leah	65	85
M	56.67	72.67
SD	6.83	9.18

Table 2

ANCOVA Summary Table (n=12)

Source	SOS	df	MS	F	Effect
Covariate	2912.038	1	2912.038	397.383	96.01%
Group	54.926	1	54.926	7.495	45.44%
Residual	65.952	9	7.328		
Total	3032.917	11			

Note: The Covariate effect size calculated was r^2=.9601 or 96.01%. The Group effect size calculated was eta^2=.4544 or 45.44%. The eta^2 result was found by dividing the Group SOS by the total SOS after the covariate effect was removed (54.926/54.926+65.952=.4544). In ANCOVA, the treatment effects are found in relation to the error scores that are created after regressing the covariate (READ) on the dependent variable (ACHIEVE). As such, the Group SOS is not divided by the Total SOS but rather by the Residual SOS after removing the covariate.

Table 3

Regression and ANOVA Summary Tables (n=12)

Source	SOS	df	MS	F	Effect
Regression					
Covariate	2912.038	1	2912.038	240.906	96.01%
Residual	120.878	10	12.088		
Total	3032.927	11			
ANOVA on Residualized Dependent Variable					
Between	17.378	1	17.378	1.6790	14.38%
Within	103.501	10	10.350		
Total	120.879	11			

Note: The r^2 effect size for the covariate (96.01%) in the regression table matches the effect size of the covariate in the ANCOVA analysis. However, the eta^2=14.38% in the ANOVA table does not match the eta^2=45.44% in the ANCOVA analysis. This is because the covariate residualized here adjusted both the dependent variable and the group membership relationship to the dependent variable.

Table 4

Classical ANOVA Summary Table (n=12)

Source	SOS	df	MS	F	Effect
Between	2380.083	1	2380.083	36.458	78.48%
Within	652.833	10	65.283		
Total	3032.917	11			

Note: Here the effect size in the classical ANOVA (eta^2=78.48%) is much larger than the effect size yielded in the ANCOVA (eta^2=45.44%). This is due to the covariate being correlated with the independent variable and removing some of the variance for which the treatment would have gotten credit had the ANCOVA not been used.

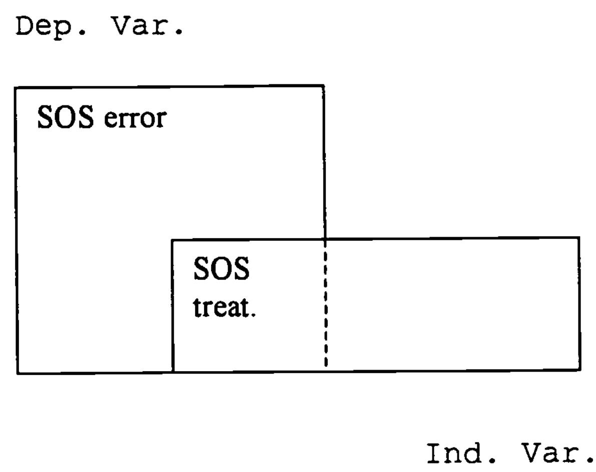


Figure 1. Venn diagram of classical ANOVA (eta^2=.25).

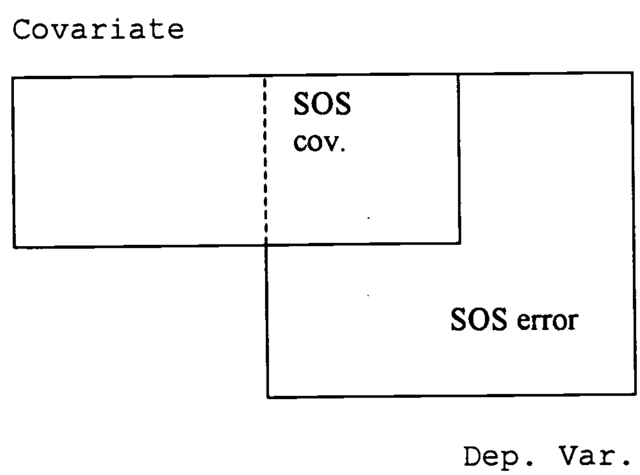


Figure 2. Venn diagram of first step in ANCOVA with covariate effect size (r^2=.25).

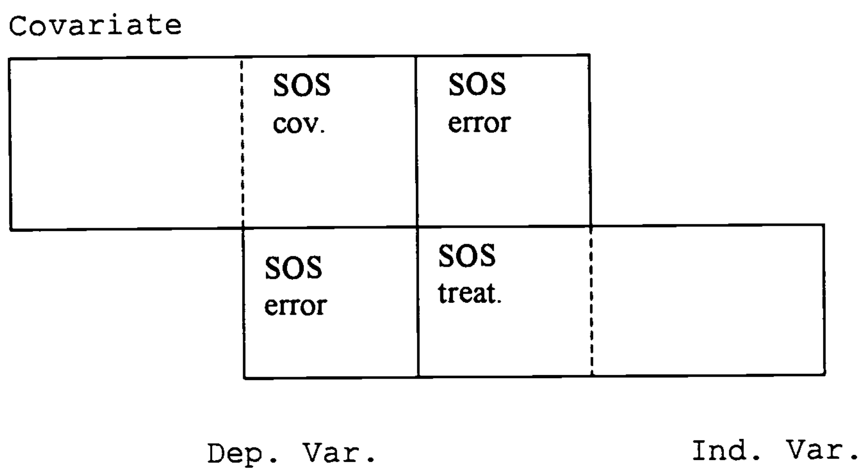


Figure 3. Venn diagram of ANCOVA analysis with enhanced treatment effect size (eta^2=.3333 of residual variance).

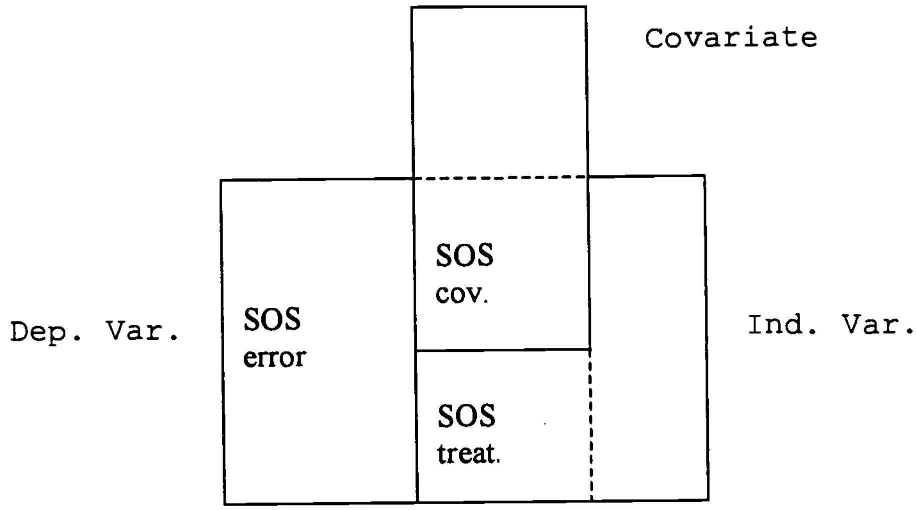
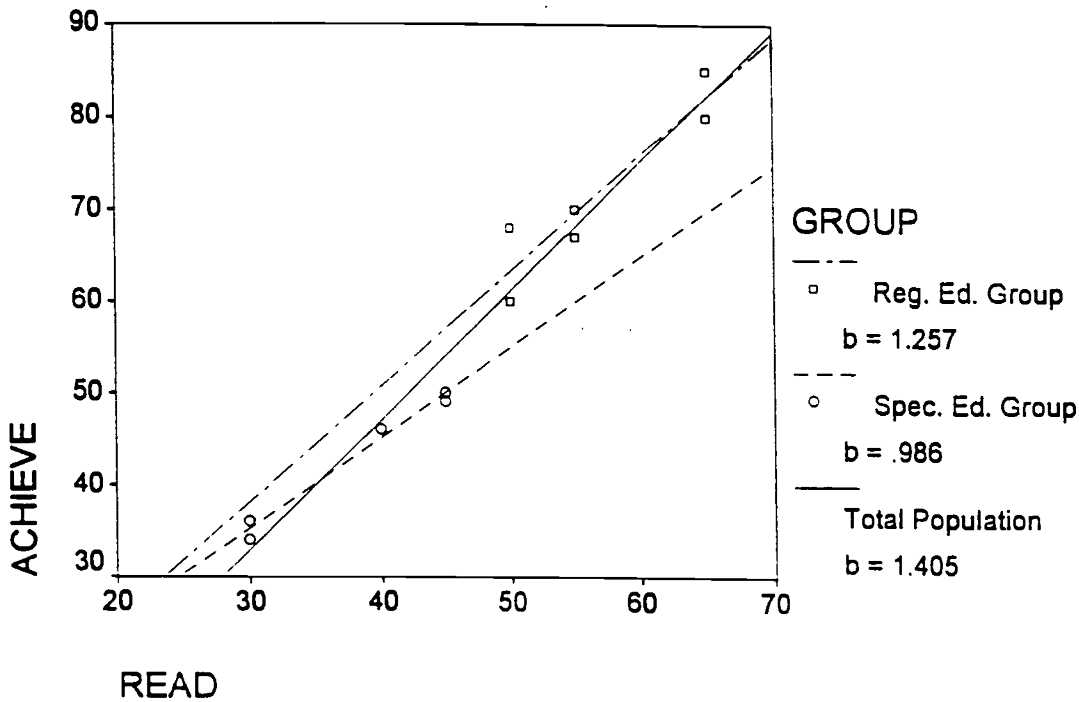


Figure 4. Venn diagram of ANCOVA analysis with correlated covariate and treatment (eta^2=.25). Treatment effect would have been 50% without a covariate correction (eta^2=.50).

Figure 5

Slopes of Subgroups and Total Sample



Appendix A

```
TITLE 'ANCOVA heuristic analysis'.
COMMENT 'n=12 see Table 1'.
GET FILE= 'a:ancova.sav'.
SET BLANKS=SYSMIS UNDEFINED=WARN PRINTBACK LISTING.
TEMPORARY.
SELECT IF (group=1).
DESCRIPTIVES
     VARIABLES=achieve read
     /FORMAT=LABELS NOINDEX
     /STATISTICS=MEAN STDDEV MIN MAX
     /SORT=MEAN (A).
TEMPORARY.
SELECT IF (group=2).
DESCRIPTIVES
     VARIABLES=achieve read
     /FORMAT=LABELS NOINDEX
     /STATISTICS=MEAN STDDEV MIN MAX
     /SORT=MEAN (A).
DESCRIPTIVES
     VARIABLES=achieve read
     /FORMAT=LABELS NOINDEX
     /STATISTICS=MEAN STDDEV MIN MAX
     /SORT=MEAN (A).
COMMENT 'determine pooled regression equation'.
REGRESSION
     /MISSING LISTWISE
     /STATISTICS COEFF OUTS R ANOVA
     /CRITERIA=PIN(.05) POUT(.10)
     /NOORIGIN
     /DEPENDENT achieve
     /METHOD=ENTER read.
COMMENT 'compute yhat using weights from all students'.
COMPUTE yhat = -9.158192+(1.405085 * read).
EXECUTE.
COMMENT 'compute residualized (error) dep var to'
COMMENT 'determine remaining dep var for ANOVA'.
COMPUTE residual = achieve - yhat.
EXECUTE.
ONEWAY
     residual BY group (1 2)
     /HARMONIC NONE
     /FORMAT NOLABELS
     /MISSING ANALYSIS.
```

Appendix A (cont.)

```
COMMENT 'show weights differ in treatment'
COMMENT 'and control groups'.
TEMPORARY.
SELECT IF (group=1).
REGRESSION
     /MISSING LISTWISE
     /STATISTICS COEFF OUTS R ANOVA
     /CRITERIA=PIN(.05) POUT(.10)
     /NOORIGIN
     /DEPENDENT achieve
     /METHOD=ENTER read.
TEMPORARY.
SELECT IF (group=2).
REGRESSION
     /MISSING LISTWISE
     /STATISTICS COEFF OUTS R ANOVA
     /CRITERIA=PIN(.05) POUT(.10)
     /NOORIGIN
     /DEPENDENT achieve
     /METHOD=ENTER read.
ANOVA
     VARIABLES=achieve
     BY group(1 2)
     WITH read
     /COVARIATES FIRST
     /MAXORDERS NONE
     /METHOD HIERARCHICAL
     /FORMAT LABELS.
ONEWAY
     achieve BY group(1 2)
     /HARMONIC NONE
     /FORMAT NOLABELS
     /MISSING ANALYSIS.
```

TM029314

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement (OERI)
Educational Resources Information Center (ERIC)

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: ANCOVA with intact groups: Don't do it!	
Author(s): ROBIN K. HENSON	
Corporate Source:	Publication Date: 11/4/98

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic/optical media, and sold through the ERIC Document Reproduction Service (EDRS) or other ERIC vendors. Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce the identified document, please CHECK ONE of the following options and sign the release below.

[X] Sample sticker to be affixed to document

Check here

Permitting microfiche (4"x 6" film), paper copy, electronic, and optical media reproduction

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

ROBIN K. HENSON

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Level 1

Sample sticker to be affixed to document [ ]

or here

Permitting reproduction in other than paper copy.

"PERMISSION TO REPRODUCE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Level 2

Sign Here, Please

Documents will be processed as indicated provided reproduction quality permits. If permission to reproduce is granted, but neither box is checked, documents will be processed at Level 1.

"I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce this document as indicated above. Reproduction from the ERIC microfiche or electronic/optical media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries."

Signature: X Robin K. Henson	Position: RES ASSOCIATE
Printed Name: Robin K. Henson	Organization: TEXAS A&M UNIVERSITY
Address: TAMU DEPT EDUC PSYC COLLEGE STATION, TX 77843-4225	Telephone Number: (409) 845-1831
	Date: 10/29/98